Legal aspects of open source intelligence

Cuijpers, C.M.K.C.

Published in:
Computer Law and Security Review

Publication date:
2013

Document Version
Early version, also known as pre-print

AUTHOR QUERY FORM

ELSEVIER	**Journal:** CLSR **Article Number:** 4795	**Please e-mail or fax your responses and any corrections to:** **E-mail:** S.J.Saxby@soton.ac.uk **Fax: +44 (0) 23 8059 3024**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	Query / Remark: Click on the Q link to find the query's location in text Please insert your reply or correction at the corresponding line in the proof
Q1	The citation "Koops et al. 2011" has been changed to match the author name/date in the reference list. Please check.
Q2	Please confirm that given names and surnames have been identified correctly.
	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

COMPUTER LAW & SECURITY REVIEW XXX (2013) 1–12

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Guest Editorial

Legal aspects of open source intelligence – Results of the VIRTUOSO project

1. Introduction[1]

Open-source intelligence involves the collection, analysis, and use of data from open sources for intelligence purposes. Using open sources for intelligence is not a new phenomenon. Already during the Second World War the US government recognized the value of openly available media sources.[2] Main interest was the acquisition, translation and analysis of foreign radio broadcasts and printed press.[3] With the coming of age of the information society, driven by the extensive private use of the Internet – unbound by time and place because of mobile smart devices – the amount of information available in open sources is enormous. Besides the obvious advantages of this rich source of information, it also entails an important downside. As Mercado puts it: "*Open sources are accessible, but they are not easy to manage. There are not only problems of scale but also of language.*" Therefore, he argues that: "*Intelligence communities require better craft to navigate the vast oceans of open sources*".[4]

Even though the Internet already provides several freely available tools such as search engines and translation programs, for law enforcement and intelligence more sophisticated means are required. As discussed by Casey, existing open source tools lack the functionality of commercial tools that are specifically designed to process network traffic as evidence.[5] And even though these commercial tools reduce the amount of time and specialized technical knowledge required to examine large quantities of network traffic, Casey stresses the fact that even these tools are from a forensic point of view, not up to standards.[6] In view of the need of specialized tools for law enforcement and intelligence, both at Community and national level,[7] initiatives are being explored to develop tools not only capable to analyse open source information, but to do this in a way compatible to legal and forensic standards.

VIRTUOSO, a project sponsored by the European Commission as part of the EC Security Research Call 2 of the 7th Framework Programme, was one of such projects aimed at designing OSINT tools.[8] As obvious as the positive contributions such tools can bring to the domain of law enforcement and intelligence, are the possible negative side-effects for fundamental freedoms and rights of citizens, whose personal information is the core value of open sources. Besides rights and obligations connected to the personal content of data, rights can also relate to intellectual ownership of data. Linked to (possible) infringements of citizens' rights is the allocation of responsibilities, anchored in legal liability regimes. Even prior to the question whether or not a framework as the one developed in VIRTUOSO is compatible with existing and applicable legal standards, the question emerges regarding desirability and acceptability of such a framework within our society. It is questionable if an operational version of the VIRTUOSO framework is reconcilable with the normative outlooks and values on which the democracies of the Member States of the EU are based.[9]

In view of these ethical and legal questions, the VIRTUOSO project contained a work package devoted to the analysis of legal and ethical implications of the development and (possible future) deployment of a VIRTUOSO framework. Such

[1] This editorial is based on research conducted in the VIRTUOSO project, in particular on Deliverable 3.2 (Koops, 2011). I want to thank the authors and reviewers of D3.2 for their valuable insights and want to thank Maurice Schellekens and Bert-Jaap Koops for their valuable contributions to both the deliverable and this editorial. Deliverable 3.2 provides a set of 20 questions that need to be answered and addressed by the partners that will perform testing and demonstration of the VIRTUOSO prototype.

[2] Glassman and Kang, p. 675.

[3] Mercado, p. 78.

[4] *Ibid.*

[5] Casey, p. 28. More on digital evidence and forensics, Casey 2011.

[6] *Ibid.* The contribution of Koops to his special issue addresses non-manipulability and auditing requirements that are associated with digital forensic quality assurance.

[7] E.g. in the Netherlands the development of IRN and iColumbo, http://columbo.nl/icolumbo/(accessed 15 July 2013).

[8] See www.virtuoso.eu. European Union Seventh Framework Programme FP7/2007–2013 under grant agreement n° SEC-GA-2009-242352. VIRTUOSO stands for: Versatile InfoRmation Toolkit for end-Users oriented Open-Sources explOitations.

[9] Deliverable 3.2 addresses ethical issues and liability issues (www.virtuoso.eu), however, this special issue is mainly devoted to the main legal barriers identified: data protection and intellectual property rights.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

research is necessary to safeguard fundamental rights and freedoms of the citizens of the European Union. The main effort of the legal and ethical work package was to develop a Privacy Impact Assessment, analyse the legal and ethical framework in open source intelligence in order to provide a list of legal and ethical requirements to be incorporated in VIRTUOSO's technical designs, an evaluation of stakeholders' interests, and an evaluation of 'code as code', to identify possible solutions to embed legal and ethical requirements into the technical designs.[10] The research findings are reported in several Deliverables available from the website www.virtuoso.eu. Building upon these findings, this special issue features three papers providing more in-depth insights into several key legal aspects relating to OSINT.

In view of the importance for other on-going EU research projects in the OSINT domain, the legal and ethical instruments developed within the VIRTUOSO project were discussed during a workshop held prior to the bi-annual TILTing perspectives Conference, in April 2013.[11] During this workshop the main legal and ethical accomplishments of the VIRTUOSO project were discussed, as well as draft versions of the contributions to this special issue. These were analysed together with several invited speakers, among which three representatives of pending EU funded projects in which similar legal and ethical questions have arisen.[12] The workshop ended with a fruitful discussion regarding the complex framework of legal and ethical constraints to be taken into consideration in the development of technologies that impact fundamental human rights and values within the intelligence domain.

This editorial will first provide some basic information regarding the type of framework developed within the VIRTUOSO project. Next, to provide the reader with some extra background knowledge, the editorial will zoom in on the two main legal constraints identified regarding the development of the VIRTUOSO framework: data protection and intellectual property rights. As on-going research projects can benefit from the more practical and organisational lessons learned within the VIRTUOSO project, this editorial features a third section in which the issues we encountered during the VIRTUOSO project, and the way in which we tried to deal with them, are shortly discussed. The editorial concludes with an introduction regarding the three papers included in this special issue. The papers provide specific and more in-depth analyses of some key legal aspects encountered in the VIRTUOSO research.

2. Main legal requirements

2.1. Introduction to the VIRTUOSO project

The aim of the VIRTUOSO project, conducted from May 2010 until June 2013, was to provide European end-users with a platform based on an open-source software solution. The platform is able to integrate advanced information acquisition and processing components – aimed at multiple kinds of open sources, available in multiple formats and media – allowing end-users to easily plug-in different software solutions (components) and, by doing so, create their own customized and modular open source information management solution. Moreover, the platform ensures greater interoperability among information and technological providers.[13]

The components developed within the VIRTUOSO project can be grouped into two classes: infrastructural components and functional components. The functional components constitute the core of the VIRTUOSO framework and include all the data processing components that will involve finding, selecting, and acquiring information from public sources and analysing it to provide relevant information useful to the end-user. The infrastructural components ensure the interactivity and the collaboration between the functional components. The main functional components of the VIRTUOSO system are: Information Gathering components (Acquisition), Information Extraction and Structuring components (Processing), Knowledge Acquisition components (knowledge management), and Decision support and visualization components. As a proof of concept of the VIRTUOSO platform and its components, a prototype has been built and demonstrated during several dissemination events. The development of a prototype, used to demonstrate VIRTUOSO functionality, made it necessary to perform a two-tier legal analysis.

2.2. Necessary distinction: research vs. end-use

In view of the analysis of ethical and legal requirements, it is essential to clarify two relevant aspects. First, this section will elaborate upon the distinction between on the one hand the development of the VIRTUOSO framework by the research consortium and, on the other hand, (possible) end-use of this framework. Second, the next section provides a brief introduction into the applicability of ethical and legal norms to open source information.

Even though VIRTUOSO is a research project and as such not directly aimed at exploitation of the research results, the prototype is developed in view of possible end-use by law enforcement and intelligence agencies. The development of the framework – including testing and demonstration of the prototype – is aimed at proving functionality and is carried out by private parties. This distinction is important, as different legal and ethical standards may apply. An example can be given from the domain of data protection legislation. Directive 95/46/EC is applicable to the researchers of the VIRTUOSO project involved with the processing of personal data. However, this directive states in article 3 (2):

> *This Directive shall not apply to the processing of personal data: - in the course of an activity which falls outside the scope of Community law, such as those provided for by Titles V and VI of the Treaty on European Union and in any case to processing operations concerning public security, defence, State security (including the economic well-being of the State when the*

[10] These correspond to the topics of the four Deliverables within WP3, available from www.virtuoso.eu.

[11] http://www.tilburguniversity.edu/research/institutes-and-research-groups/tilt/events/tilting-perspectives-2013/program/ (accessed 15 July 2013).

[12] Representatives from the Sapient project and the Caper project were present, http://www.sapientproject.eu/ and http://www.fp7-caper.eu/(accessed 15 July 2013).

[13] www.virtuoso.eu (accessed 15 July 2013).

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

processing operation relates to State security matters) and the activities of the State in areas of criminal law.

In this respect, a special Framework Decision is effective regarding the processing of personal data in the domain of police and judicial cooperation in criminal matters.[14] Even though the Directive and the Framework Decision are similar, important differences relate to specific exceptions and legitimate processing grounds that are reserved for public authorities in the field of law enforcement and intelligence, creating more leeway for them than is the case regarding private parties.

Because of differences in applicable legal regimes, it was necessary to perform a two-tiered legal and ethical analysis within the VIRTUOSO project. On the one hand the legal assessment of the prototype, as embodiment of the VIRTUOSO platform and relevant components. In demonstrating VIRTUOSO functionality the researchers involved needed to perform acts with legal implications, such as the processing of personal and/or copyright protected data.

The second tier of the legal and ethical analysis concerned the foreseen end-use by law enforcement and intelligence agencies. As explained, actual end-use is outside the scope of the VIRTUOSO research project. Still, for the development of the VIRTUOSO framework, it is necessary from a human rights and ethical perspective to reflect upon the legitimacy of the foreseeable end-use of the framework. This reflection relates to the overall legitimacy of developing a framework such as VIRTUOSO, and to some extent also on the expected legal and ethical risks associated with the foreseeable actual end-use. From this perspective, the design of VIRTUOSO should encompass necessary safeguards and use limitations to minimise such risks.

Even though actual end-users have a responsibility of their own for ethical and legal compliance, it is important to recognise that it is questionable whether all responsibility for a proper functioning and use of VIRTUOSO can be ascribed to the end-users. Some responsibility for a proper functioning of the VIRTUOSO framework in practice also lies with the developers of the platform and individual components. Therefore, the second tier of research addressed legal and ethical questions regarding developers' responsibility for compliance with legal and ethical standards by probable, or possible, end-use of the developed framework.

Interesting to note is that the ethical and legal analysis taught that more restrictions seem to apply to the researchers within the VIRTUOSO project, than is the case with the actual end-users. Actual end-use might have a huge impact on society; however, foreseen end-users quite often can call upon special authorities and competencies in the field of law enforcement and/or intelligence. Which, in contrast, do not exist for the private parties developing and testing the VIRTUOSO framework and prototype.

2.3. Open does not equal free

Open-source intelligence involves the collection, analysis, and use of data from open sources for intelligence purposes. As it turned out, the terminology of *open* gave rise to a common misunderstanding amongst most technical partners involved in the VIRTUOSO research. This misunderstanding became clear because of a recurrent argument in the initial stages of the research: *"We only use **open** sources, so we have nothing to do with the existing legal framework."* It took some convincing, and academically substantiated explanations, to demonstrate that the fact that data are openly available does not mean that they can be processed without regard to legal and ethical standards. Put in other words, the mere fact that data are publicly available does not imply an absence of restrictions to researching them. From the legal and ethical quick-scan performed during the first six months of the VIRTUOSO project, it became clear that data protection and copyrights create specific restrictions in view of the development and use of the VIRTUOSO framework and prototype.

Another issue to deal with in relation to open sources concerned the lack of clarity and uniformity regarding the scope of the notion "open source".[15] This closely relates to one of the most important lessons learned within the VIRTUOSO project, the need for a common language as further described in Section 3.2. In view of open sources it was decided amongst the VIRTUOSO research partners to define these as: "Sources that are freely available to the public, i.e., information sources that have no access or other types of restrictions nor payment requirements".

2.4. Privacy and data protection

2.4.1. Introduction

In European legislation, privacy is embedded as a fundamental human right in article 8 of the European Convention on Human Rights and Fundamental Freedoms (ECHR). The Charter of Fundamental Rights of the European Union, besides the right to privacy (art. 7), also contains a fundamental right regarding the protection of personal data (art. 8). Moreover, data protection is strictly regulated in Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

As explained above, the legal analysis within VIRTUOSO has its focus on the development of the VIRTUOSO framework, where end-use scenarios are only addressed in order to asses to what extent the development of VIRTUOSO must anticipate foreseen end-use, in order to incorporate requirements and safeguards into the VIRTUOSO design. The core of the legal analysis is aimed at the acts performed by partners of the researchers consortium in order to develop the prototype and prove its functionality. To align the development with legal and ethical standards a comprehensible list of requirements was deducted from relevant existing legislation. The list formed the basis for the legal compliance assessment, and was used as a starting point to investigate how – if possible – to incorporate legal standards into the VIRTUOSO design.

Below the checklists for privacy and intellectual property rights are presented, as these are useful for the development of other tools that process open source information. In general, within these lists references to VIRTUOSO can be

[14] Council Framework Decision 2008/977/JHA, OJ L350, 30/12/2008, p. 60-71.

[15] See also the contribution of Koops in this special issue.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

replaced by any other OSINT technology. The second phase of the research, actually applying the list to the VIRTUOSO project, will not be described in this editorial as it bears less relevance for a broader audience. The main legal barriers in respect of the development and proof of functionality of the VIRTUOSO prototype will be discussed, as these probably impact all kinds of OSINT tools.

It is important to note that the data protection checklist was developed on the basis of the existing EU legal framework. In January 2012 proposals were published for a major EU data protection reform, replacing Directive 95/46/EC with a regulation and Framework Decision 2008/977/JHA with a directive.[16] The importance of the proposed reform for the development of new technologies that might impact upon privacy and data protection will be addressed in Section 3.3.

2.4.2. *Privacy checklist*

There is a relevant distinction between privacy and data protection, as both rights have their own legal framework. The right to privacy as laid down in art. 8 of the ECHR, holds three steps to assure whether or not certain actions infringe upon this right to be free from interference from others. On the basis of this article, the following questions must be answered in order to assess privacy in relation to the VIRTUOSO framework:

1. Does (the use of) VIRTUOSO interfere with the privacy of citizens?
2. If so, does (the use of) VIRTUOSO have a legal basis?
3. Is one of the grounds mentioned in art. 8 ECHR applicable (e.g. national security and the prevention or detection of crime)?
4. Is VIRTUOSO necessary in a democratic society? This entails the following sub- questions:

a. To what extent does VIRTUOSO actually contribute to its goal, and are there less invasive alternatives to reaching this goal (subsidiarity principle[17])?
b. Is the privacy infringement outweighed by the importance of the goal, and what guarantees does VIRTUOSO offer to diminish the privacy infringement (proportionality principle)?

The privacy test mainly concerns actual end-use, as the impact on privacy of the research as such is limited. The purposes of processing personal data are restricted to demonstrating the prototype while no actual decision making on the basis of such data is involved. It is however important to keep the test of art. 8 ECHR in mind in the development of the VIRTUOSO framework, as the fourth question illustrates the importance of designing the framework in such a way that end-use is at least capable to comply with the principles of proportionality and subsidiarity.

2.4.3. *Data protection checklist*

For data protection a similar checklist was developed, incorporating all relevant requirements that follow from Directive 95/46/EC. In view of this checklist it is important to stress the fact that being able to answer the questions is not sufficient to comply with the legal requirements. Proper, clear and complete documentation regarding data processing being carried out should be provided for. In view of several obligations stemming from Directive 95/46/EC this is mandatory, e.g. purpose specification, for other obligations documentation is simply necessary to substantiate claims and choices being made regarding the technical design. Especially in view of the substantiation to withstand the proportionality and subsidiarity test of the right to privacy as discussed in the previous section.

Without providing an in-depth analysis of the legal framework regarding data protection, which is not necessary to be able to understand the list of requirements, three key concepts of the framework will be introduced. Data protection legislation is all about "*fair and lawful processing of personal data*". Personal data is any data relating to an identified or identifiable natural person. Processing concerns every handling of personal data, from the creation up to destruction of such data. If data processing is fair and lawful depends on the question whether all requirements, contained in the checklist below, have been complied with.

1. Does (the use of) VIRTUOSO involve the processing of personal data, i.e., data relating to directly or indirectly identifiable individuals? If so, the following questions must be answered.
2. Who is the data controller, i.e., the one who factually determines the purposes and means of the processing of personal data? The data controller has to answer the following questions and take the necessary measures to safeguard compliance with data protection legislation mentioned in the following questions.
3. Is there a legitimate purpose (or more) to collect and process personal data?
   - Is that purpose (or are those purposes) clearly specified?
   - Is the way in which the data are subsequently processed compatible with the purpose or purposes for which they were collected?
4. If not, is there a legitimate reason why data can be processed in a way that is incompatible with the purpose or purposes for which they were collected on the basis of any of the exceptions?
5. Is there a legal ground (provided for in art. 7 Directive 95/46/EC) that legitimates the processing (including the collection)? I.e., is there (informed and explicit) consent of data subjects, a legal obligation, or a substantial interest that outweighs the privacy interest of the data subject?
5. Are measures taken to ensure that the personal data are correct, accurate, adequate, relevant and not excessive, in view of the processing purpose(s)? In other words, are no more data being processed then strictly necessary in view of the specified purposes? This closely relates to specifying the purposes of (the use of) a system based on the VIRTUOSO platform.

[16] Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (Text with EEA relevance) {SEC(2012) 72 final} {SEC(2012) 73 final}. Brussels, 25.1.2012, COM(2012) 11 final 2012/0011 (COD).

[17] Also known as the principle of least onerous means.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

6. Are data retained longer than necessary? Personal data has to be deleted and disposed of once its value or utility has come to an end.
7. Are data properly secured, using appropriate security measures that are state-of-the-art and cost-effective?
8. Are information obligations met? I.e., are data subjects informed of the processing, and can they exercise access and correction rights? A proper means to disseminate such information is the VIRTUOSO website including a contact address that data subjects can approach with a request for information and, where relevant, correction of their personal data that are being processed.
9. Are notification obligations met? I.e. the processing of personal data must be notified to the Data Protection Authority of the country or countries in which the data are processed.
10. Are sensitive data being processed, for example personal data that relate to ethnicity, religion, sexual life or health? When images and videos are being processed, chances are high that the severe regime regarding sensitive data applies, as pictures and video often reveal sensitive data such as religion (e.g., headscarf), ethnicity, and sometimes even health.
11. Is the processing of sensitive data covered by any of the general or specific exemptions from the prohibition from processing such data?
12. Are data being transferred to a country outside the European Union? If so, does that country have an adequate level of data protection?

2.4.4. *Main data protection constraints*

The above list demonstrates a multitude of requirements that must be met, and where possible be anchored in the VIRTUOSO design. Information obligations, security standards and data subject rights could be embedded into the overall VIRTUOSO design by technical, organisational and procedural measures. The assessment of how to do this will be quite a challenge. Not only from a technical perspective, but also from a social and economic viewpoint; i.e. which mechanism will be trusted and accepted by society, which measures will be most effective and which measures to choose on the basis of a cost-benefit analysis. However, in principle the implementation of such requirements does not directly constrain VIRTUOSO functionality. This is however the case for the requirement of purpose specification and limitation, as will be explained below. The requirement of a legitimate ground even poses a severe problem to process open source data including sensitive data to demonstrate the functionality of the VIRTUOSO framework and prototype.

Broadly speaking, the purpose of processing personal data within the VIRTUOSO research project is to test the prototype and to demonstrate its capabilities and added-value with regard to the needs expressed by end-users. As it is unlikely that a lawful exception is available for processing data in a way that is incompatible with this primary purpose, the partners involved in testing must guarantee that their activities are necessary for testing the prototype and determining its functionality. Since the purpose of the VIRTUOSO system is open-source intelligence, in which – at the core of the system – large amounts of data from open sources are collected and subsequently analysed, it is not required to limit the collection of personal data ex ante, in processing phases which only concern fully automated collection and transformation of data into a standardised format. However, the principle of fair and lawful processing does require that personal data are only processed to the extent really necessary. The need to process actual data therefore needs to be properly substantiated and justified. Where and if possible, actual names should – as soon as possible – be anonymised or pseudonymised in an irreversible way (e.g., using one- way hash functions). For personal data that are not anonymised or irreversibly pseudonymised, it must be ensured that they are correct and accurate. This is in line with the general task and professional standards of data analysts, but it underlines the importance of checking very carefully the accuracy and correctness of data that are found in open sources, which have a non-negligible risk of being inaccurate or incorrect. Personal data may be stored only for as long as is strictly necessary for testing the prototype. As soon as the prototype has been successfully tested, these data will have to be deleted permanently and securely. In any event, all data collected should be deleted before the end of the VIRTUOSO project. If part of the prototypes function is learning capability, for which it is essential to hold on to data longer than in view of the demonstration of other functionalities, this must be specified in the purpose in order to legitimise the storage of data for a longer period of time in order to demonstrate the learning capability functionality.

The fact that personal data are available online does not imply that users have given consent to use these data for other purposes than those for which they were published online. Moreover, it is likely that a significant proportion of the personal data available in open sources will not have been published online by the data subjects themselves, but by other parties. So, it is not possible to base the processing of personal data available in open sources on consent. A legitimate ground cannot be found either in a contract with data subjects or a legal obligation that VIRTUOSO is under to process the personal data. For the processing of personal data this does not have to be problematic. According to art. 7(f) of Directive 95/46/EC, personal data processing is allowed when VIRTUOSO researchers can demonstrate they have a substantial interest in such processing that outweighs the privacy interest of the data subjects whose data are being collected and processed. This underlines the importance of VIRTUOSO explicitly substantiating the purpose and necessity of developing and testing the prototype. In processing sensitive personal data, the difficulty with the requirement of legitimate grounds does come into play. The VIRTUOSO system is not capable, at least not in the initial collection phase, to distinguish between types of data. As VIRTUOSO includes images and videos being processed, chances are high that the severe regime regarding sensitive personal data applies as images and videos are likely to reveal religion, ethnicity or health issues. Therefore, it is quite certain that within VIRTUOSO sensitive data will be processed. Directive 95/46/EC has a strict regime regarding processing of sensitive data. Article 8 of Directive 95/46/EC states:

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life.

There are some exceptions in the second section of article 8, when: the data subject has given explicit consent; processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law; processing is necessary to protect the vital interests of the data subject or of another person; processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim; processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims. The third section of article 8 concerns an exception for processing in relation to preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, while section 4 allows Member States to, for reasons of substantial public interest, lay down additional exemptions.

If sensitive data are processed, it is questionable whether any of these legal exemptions can apply to the researchers involved in VIRTUOSO. The only possible exemption is when *"the processing relates to data which are manifestly made public by the data subject"*. As it will be difficult to establish – without any doubt – that a photograph has been uploaded by a data subject herself and does not include other identifiable individuals, this ground doesn't seem to offer a sound solution either. Perhaps other exemption grounds can be found in the national law of the country where the data are processed, e.g. exemptions for scientific research, as the Data Protection Directive allows Member States to stipulate additional exemptions in their national law. However, this would entail the use of open source data in order to demonstrate VIRTUOSO functionality to be confined to data solely governed by jurisdictions encompassing such an exemption.

2.5. *Intellectual property*

2.5.1. *Introduction*

Besides restrictions stemming from privacy and data protection, data in open sources can be subject to intellectual property rights (hereinafter IPRs). For OSINT, copyright and database rights are the most relevant IPRs. Much information that is available in open sources is protected by these IPRs. Importantly, it cannot be assumed that a rights holder has waived his rights, merely because his work is available to the public for free or a copyright notice is missing. The same holds true if content has not been protected by technical means, the rights holder does not restrict access to the work, the rights holder does not indicate that he retains, exercises or enforces his right, or because of some other reason that does not include a waiver of rights. If a work is protected by copyright or database right, acts relevant under these rights (i.e. reproduction or extraction) can only be performed with a license from the rights holder or where an exemption to an exclusive right is applicable.

2.5.2. *VIRTUOSO as a mere conduit or service provider*

If VIRTUOSO is exploited beyond the phase of mere demonstration of its functionality, two scenarios are possible regarding VIRTUOSO's role in relation to end-users. First, VIRTUOSO can merely provide the platform. Second, VIRTUOSO can take the role of service provider. In view of liability for the infringement of copyrights and database rights, determination of the exact role is relevant as different liability regimes apply.[18] The platform provider that offers only the architecture, including technical components, but without actually functioning as a service provider for hosting services, can be considered to be a mere conduit provider. Such a provider is not able to intervene in the data transfers that take place over the platform. This would make the platform provider largely exempt from liability, provided that he has no actual control over any processing of data or storage of data in repositories. As a service provider, VIRTUOSO would be offering hosting services, maintenance services or automated translation modules or other automated data enrichment services. In this scenario there would be no direct involvement with copyrighted information; hence the service provider as such is unlikely to be found to be infringing copyrights. Unlike a platform provider, the service provider is however able to intervene in the storage or processing of data. Even though not infringing directly, the capacity to intervene in third party infringement is the reason that the law imposes certain duties of care upon a hosting provider. These duties of care are likely to encompass stopping clear infringements by end-users that the provider knows of, and a duty not to incite others to infringement. End-users who deal directly with protected information have to closely scrutinize their activities and the data they are dealing with, as they are likely to be directly liable for any violations of intellectual property rights.

2.5.3. *IPRs: the main issues*

The following questions require an answer in order to assess the legality of the use of open source data within the context of OSINT:

1. Are the collected data protected by Intellectual Property Rights (IPRs)? If so,
2. Is the use of the data relevant under the IPRs? If so,
3. Is the use of data exempt under some statutory limitation of the exclusive rights? If not,
4. Has the rights holder given an implicit license by placing the material in open sources? If not,
5. Can an explicit license be obtained?

2.5.4. *Copyright*

In Europe, works are protected under copyright if they are original creations of their authors. According to the EU Court

[18] Art. 12–14 of Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, Official Journal L 178, 17/07/2000 P. 0001–0016.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

of Justice in its Infopaq decision this means that the work is an own intellectual creation of its maker.[19] Articles written by (Internet) journalists meet almost without exception this requirement. Copyright protection comes into existence 'automatically' as soon as the work is made available. No registration of a copyright is required. Since the threshold for protection is relatively low and protection comes automatically in existence, many texts, images, audio recordings and movies comprised in web pages will be protected under copyright. Copyright grants the maker of a work an exclusive right to reproduction, to communicate the work to the public and an exclusive right for adaptation and translation.[20] The applicability of an intellectual property right does not mean that all acts that are relevant under the legal regime are not allowed. The use of protected materials can be justified by a license or a statutory exception, as will be elaborated below.

2.5.5. *Data base right*

A database is defined as a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.[21] Many archives of e-mail lists qualify as databases. Often databases are made available on the public part of the Internet. These databases may receive protection under copyright and under sui generis database rights. The protection under copyright concerns the structure of the database, i.e. the selection and arrangement of its elements. This type of protection for databases is relatively weak. However, sui generis database protection extends to the contents of a database, i.e. the entirety of the elements contained in the database. The contents are protected if a substantial investment has been made in the collection, verification or presentation of the contents of a database. The EUCJ has heightened the threshold of protection somewhat by deciding that investment in the collection of data does not include investments done in the actual creation of the data.[22] Only the investment in the collection of existing data 'counts'. Nevertheless, many of the databases in open sources will have attracted database protection. Just as in copyright, database protection comes 'automatically' into existence once the database has been completed.

The database right grants the maker of a database an exclusive right for extraction and reutilisation. These rights are comparable to the reproduction right and the right of communication to the public in copyright law. The database right does not encompass a right of adaptation or translation. Similar to copyrights, the use of protected database contents can be justified on the basis of both explicit and implicit licensing, or on the basis of an exemption.

2.5.6. *Licenses*

If a rights holder gives permission to use protected materials in the context of VIRTUOSO, this is obviously allowed. The rights holder's permission may come in one of several forms: as an implicit or an explicit license, in standard form or specifically negotiated, machine readable or requiring human interaction, against compensation or for free. A rights holder who has placed his work on the Internet may indicate what uses of the work he allows and under what terms. The terms may typically be found under a hyperlink entitled 'terms of use' often located at the bottom of a webpage. The terms may constitute an offer that the Internet user accepts by using the website or data or acceptance may require some interaction between the user and the (computer of the) rights holder specifically related to the license. The latter may in particular be the case if the license is not for free. Of particular interest are licenses according to the Creative Commons format.[23] These licenses are standardised and may allow use for free. They allow for non-commercial use of the works they are attached to, but may disallow or restrict adaptations of the works.

Where no explicit license is present, sometimes an implicit license can be deduced from the way in which a work is made available. If for example a rights holder places his work on the public part of the Internet under a 'Download' button it can be assumed that actual downloading is not infringing his copyright, even though downloading amounts to the making of a reproduction. In many cases however, it is difficult to assess to what uses of a work an implicit license extends, and national courts have handed down diverging judgements. According to a German court Google Image search may display thumbnails of images in its hit lists. A photographer placing her images on the Internet was found to have given an implicit license for use of thumbnails of her images by a search engine. According to a Belgian court Google News may not display the titles and the first few lines of news messages found on other news sites when linking to these other news sites.[24] The news sites were not thought to have given an implicit license.

There certainly is an argument to be made that publishing material on the web involves an implicit license for some use of such materials. Technologies as the ones being developed in VIRTUOSO make information better accessible and searchable and given the data volumes present on the Internet they may be necessary to make use of the Internet. There is however no court that has ever subscribed to this view in the context of OSINT. Moreover, some courts require for reliance on an implicit license that the rights holder must have placed the information on the Internet herself, which in practice may be difficult to verify when processing vast amounts of data.[25]

[19] CJEU 16 July 2009, case C-5/08, Infopaq International/Danske Dagblades Forening, online available at http://www.curia.eu (Accessed 15 July 2013).

[20] Articles 2 and 3 of Directive 2001/29/EC and articles 8 and 12 of the Berne Convention. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, pp. 10–19. Berne Convention available from www.wipo.int (Accessed 15 July 2013).

[21] Art. 1 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Official Journal L 077, 27/03/1996 P. 0020 - 0028.

[22] ECJ 9 November 2004, Case C-203/02, The British Horseracing Board Ltd and Others v William Hill Organization Ltd.

[23] http://creativecommons.org/licenses/

[24] For a discussion of these cases P.B. Hugenholtz and M.R.F Senftleben, *Fair use in Europe. In Search of Flexibilities*, Amsterdam, November 2011, p. 17.

[25] *Ibid.*

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

Hence, it is under the current EU legal regime risky to rely on an implicit license.

2.5.7. Statutory exemptions

The use of protected content to which a third party holds the rights may be allowed if the use in question is covered by a statutory exemption. In relation to VIRTUOSO, relevant copyright exemptions are the exception for transient copying and the research exception and the exception for public security. The research exemption may be relevant for Virtuosos, since this project is still in the development stage and the testing of a prototype may qualify as 'research'. Article 5, section 3 (a) of Directive 2001/29/EC allows for an exemption from the reproduction right and the right of communication to the public in case of scientific research. EU Member States have used the possibility offered by the Directive to create a research exemption in their national Copyright Acts to varying degrees; in other words, there is little actual harmonisation. The French Copyright Act does for example not contain a research exemption. It is thus relevant to determine in what Member States relevant acts under intellectual property laws take place and evaluate these acts under the local IPR regime. The research exemption as defined in the directive does not extend to the adaptation right.

Art. 5.1 of Directive 2001/29/EC reads as follows:

> *Temporary acts of reproduction referred to in Article 2, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article 2.*

Member States of the EU must implement this exemption from the reproduction right. Hence, there is strong harmonisation within the EU on this point. Technical copies in CPU registers or RAM memory that are made in the context of automated checking of data (that constitute works) could be covered by this exemption. Bottleneck will be that such transient copying will need to take place in the context of a lawful use of a work. This implies some dependence on the other instruments for justifying uses of works.

In view of Database protection, article 9 of the Database Directive provides for several exemptions amongst which those for public security and scientific research:

> *Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:*
>
> *(a) in the case of extraction for private purposes of the contents of a non-electronic database;*
>
> *(b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;*
>
> *(c) in the case of extraction and/or re-utilization for the purposes of public security or an administrative or judicial procedure., amongst which public security and scientific research.*[26]

Just as is the case with copyright, Member States need not implement such exemptions in national legislation. Therefore, differences may exist depending on applicable national law.

2.5.8. Applicable law and national exemptions

Given that the law in the different Member States is fragmented, it may be necessary to determine the law of which Member State is applicable. The law applicable to infringement of copyright or a database right is the law of the country in which the rights holder seeks to have his rights protected. For example, the law applicable to an injunction of a copyright infringement is the law of the country where the (alleged) infringement takes place and rights holder seeks to have it stopped. Application of this law is governed by the assimilation principle in copyright law and by the reciprocity principle in database rights' law. The assimilation principle means that foreign authors are treated at least as well as national authors. The reciprocity principle means that the database rights only apply to databases of rights holders from third countries to the extent that those third countries offer comparable protection to databases as applies in the EU.

2.5.9. Conclusion

OSINT necessarily involves reproduction and adaptations (e.g. conversion to other formats), acts relevant under copyright protection. It might also involve extraction under database protection, but the risks in relation to copyrights are more severe. Under copyright, the only certain way to legitimize reproductions and adaptations is through explicit licenses. It may not be practical, however, to apply for explicit licenses for all sources, nor to limit searches to sources that have (automatically identifiable) explicit licenses on their websites. There is an argument to be made that publishing material on the web involves an implicit license for adaptations and translations, but this argument has not been tested in court. Therefore, processing material from open sources without an explicit license, under the assumption that an implicit license is given by the rights holder, involves a certain liability risk of copyright infringement.

To conclude, besides legal constraints stemming directly from intellectual property rights, site owners can also revert to organisational, procedural and/or technical access constraints. In some cases, web contents are made off-limits for bots with the help of a robots.txt file or in the meta-information of a website. The contribution of Schellekens to this special issue analyses the use of this specific access constraints and the questions it raises from a legal perspective.

2.6. Conclusion

The analysis of both the legal domain of privacy and data protection, and the domain of intellectual property rights,

[26] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal L 077, 27/03/1996 P. 0020–0028.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

illustrate the importance – and dependence – on national implementation legislation. Even though some international regulatory guidelines are available, specific allowances, prohibitions and exceptions mainly stem from national legislation. The difficulty in this respect lies with the fact that national legislation can differ heavily. For applications such as the VIRTUOSO framework this is a major concern, as the acts performed on the Internet are hard to confine to the boundaries of one specific country. Therefore, at a general level of technology assessment, legal research seems to be limited to rather general guidelines and conclusions, which bear more significance when applied to actual concrete and nationally oriented cases. The dependency and influence of national law in the specific domain of procedural law is illustrated by the contribution of Koops in this special issue.

3. Lessons learned

3.1. Introduction

Besides the legal and ethical requirements defined within the VIRTUOSO project that benefit the development of OSINT tools – which requirements can be used in the much wider domain of developing technologies with an impact on human rights – the VIRTUOSO project also was a good learning experience regarding more practical and organisational issues of multidisciplinary research. Future projects involving technological developments that may impact ethical values and legal norms, might benefit from the lessons learned within VIRTUOSO. The next sections will briefly address two important lessons learned.

3.2. Developing a common language

Even though all project partners were capable to communicate in English, it was clear from the start of the project that despite using the same words, we did not speak the same language. Already during the kick-off meeting it became obvious that researchers within different scientific domains, all speak their own jargon. Especially for the legal and ethical researchers the lack of understanding of the technical jargon, and the need to have proper and clear definitions, turned out to be a difficulty that needed to be addressed at an early stage in the project. To give just one example, there was great inconsistency in the terms used to explain the technical design of VIRTUOSO. Words like 'tool', 'toolkit', 'component', 'infrastructure', 'platform', 'factory', 'prototype', and 'demonstrator' were used in order to describe the overall – or parts of – the initial designs of VIRTUOSO. Although largely irrelevant to the technical partners, these inconsistencies created problems in view of properly assessing legal and ethical requirements.

To deal with this issue, a wiki and taxonomy were created within the private space of the VIRTUOSO website. This interactive setting allowed technical and legal/ethical partners to provide, question, discuss and amend definitions of key concepts used within the project. This way a simplistic, but manageable set of concepts was agreed upon to denote the different technical outputs of the VIRTUOSO project. This clarity in terminology made it possible to gear the legal and ethical research towards different 'layers' of the technical design.

The wiki and taxonomy were also relevant in view of the previously discussed misconceptions of "open" meaning "free of morals, rights and obligations" and another commonly heard argument: "we are not doing anything, as we are not end-users". The wiki mainly was used to give more guidance and explanation regarding the meaning and applicability of legal and ethical standards, as a point of reference for the engineers. To improve common understanding, several workshops were organised in which the legal and ethical scholars gave lectures in privacy and intellectual property rights. Vice versa, by simplifying the technical designs in flow charts and pictures, the technicians were able to make the lawyers understand at least the basics of the technical VIRTUOSO designs and the functionalities to be embedded and employed in the VIRTUOSO framework, necessary in view of the legal and ethical assessment.

3.3. Chicken-and-egg-problem

Finding a common terminology was an important step in the creation of mutual understanding. However, it did not solve one important problem related to the multidisciplinary nature of the research, which can most easily be referred to as the 'chicken-and-egg problem'. In order to embed legal and ethical requirements into a technical design, these requirements need to be clear from the outset. However, technical and legal work packages both commenced in the first month of the VIRTUOSO project. Another problem concerned the fact that in order to give a proper legal and ethical analysis, it needs to be clear, at least to some extent, what the specifications and functionalities of the technical design are. However, the technical developers wanted such decisions at least in part to be based on legal and ethical guidelines.

To cope with this problem, a legal and ethical quick-scan was performed during the first months of the project, so the technical partners at least had something to work with. Throughout the first two years of the project this quick scan was further developed to provide concise guidelines for the actual prototype. In order to actually embed legal and ethical guidelines into technical designs an on-going process of mutual reflection, both from the technical as the legal/ethical perspective was necessary. This can be denoted as a 'zig zag strategy' which involves successive accretions of scientists from both the legal and technical domain, in order to build up and further the knowledge in both domains and to intertwine this knowledge within the technical design.

3.4. Connecting practice to theory, relevance of privacy by design

3.4.1. Introduction

Identifying legal and ethical requirements is just a first step. More interesting is the creative process to find ways to incorporate these requirements into the technical VIRTUOSO designs. The aim of the technical designs must be to safeguard that the functionality of the end product(s) is as minimally

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

invasive for ethical standards and legal rights, while ensuring it can serve its intended purposes of open-source intelligence. As became clear during the VIRTUOSO research, to incorporate legal and ethical standards into the designs takes a certain mind-set, involving awareness and acknowledgement of the need to address these aspects right from the start of the project. In view of privacy, the proposed Regulation provides a severe incentive for such a mind-set, as it explicitly prescribes the principles of data protection by design and privacy by default.

The need to embrace privacy by design definitely is an important second lesson learned within the VIRTUOSO project. The contribution of Koops, Hoepman and Leenes gives more in-depth insight into the possibilities Privacy by design has to offer by describing various strategies and two specific potential solutions to incorporate normative and legal requirements into the VIRTUOSO design. Here, because of its importance, some background will be given to the concept of privacy by design. It must be stressed that incorporating other legal and ethical requirements besides privacy is just as important. However, the proposed mandatory nature of privacy by design provides an excellent explanation and incentive for engineers to actually adapt the required mind-set.

3.4.2. *Privacy by design and default*

The proposed Regulation defines privacy by design in article 23 as:

> *Having regard to the state of the art and the cost of implementation, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures and procedures in such a way that the processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject.*

Besides the principle of privacy by design, the Regulation even goes one step further by requiring privacy by default, meaning:

> *The controller shall implement mechanisms for ensuring that, by default, only those personal data are processed which are necessary for each specific purpose of the processing and are especially not collected or retained beyond the minimum necessary for those purposes, both in terms of the amount of the data and the time of their storage. In particular, those mechanisms shall ensure that by default personal data are not made accessible to an indefinite number of individuals.*

In order to assure compliance, the Regulation provides for severe penalties. Article 79 (6) (e) states:

> *The supervisory authority shall impose a fine up to 1 000 000 EUR or, in case of an enterprise up to 2 % of its annual worldwide turnover, to anyone who, intentionally or negligently:*
>
> *(e) does not adopt internal policies or does not implement appropriate measures for ensuring and demonstrating compliance pursuant to Articles 22, 23 and 30.*

In view of the above, not only for ethical reasons, but also to minimize liability risks, developers of OSINT tools such as VIRTUOSO have an obligation to limit the potential of abuse of the system, both by legitimate and by illegitimate end-users. The scope of such obligations depends on the wording and interpretation of the principle of data protection by design, as this principle is directed to controllers. However, at least in the pilot phase, developers will also be controllers. Moreover, besides the liability argument, from a market perspective end-users will demand tools and products capable to comply with data protection by design. Where possible, constraints must be built into the design of the system in order to safeguard legal and ethical standards. These built-in constraints can be technical or organizational in nature, or a combination of both. Within the VIRTUOSO project this is a general requirement that both individual partners developing components and the overall platform development partners have to take into account.

Mainly for developers the starting point of privacy by design is an important lesson. They may not start from the idea that as much functionality as possible should be developed to maximize data collection and knowledge production as a goal in itself. Instead, the starting point should be to create an optimal combination of a) functionality that benefits open-source intelligence and b) protection of legal rights and ethical values. In some cases, this could (and should) lead to closing off certain functionality if that can be particularly harmful to individuals whose privacy is at issue or to intellectual property rights holders. This can be achieved by incorporating technical, organizational and procedural measures into the system. Such measures can be used to minimize function creep, meaning the use of the product(s) for purposes other than their originally intended purposes; to minimize misuse of the product(s), e.g., by measures to enable and ensure restricted access, authorization, logging of use; to minimize the processing of personal data, e.g., by anonymising or deleting personal data as soon as possible, and by securing personal data against leaking, e.g. by state-of-the-art encryption; to minimize the risk of violation of intellectual property rights, e.g., by automatically complying with rights holders' stipulations in robots.txt files or website meta-information.

3.4.3. *Dissemination*

To close the development cycle, it is essential to make the measures taken to accommodate ethical and legal standards available to the public at large. In other words, it must be articulated how e.g. privacy by design efforts have been embraced in the development of a specific OSINT tool and how they have affected the end-result. Providing proper information is especially important when products are marketed and distributed to end-users. Information to end-users about responsible use, including explanations regarding legal and ethical requirements associated with the use of the product, should be provided for in leaflets and product descriptions, but also in the general terms and conditions and/or contracts with end-users. To minimize liability risks a disclaimer should be included for those cases in which end-users disregard legal and ethical requirements.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

4. Contributions to this special issue

This special issue of the Computer Law & Security Review features three contributions, written by several researchers involved in the VIRTUOSO project. As explained above, the contributions aim to provide more in-depth insights into specific aspects of the more generic output provided for in the VIRTUOSO deliverables.[27]

The first contribution by Koops concerns procedural-issues in police investigations in Internet open sources. In view of the identified legal and ethical restrictions regarding the use of open sources, specifically in relation to the domain of law enforcement, this paper investigates one area of legal constraints: criminal-procedure law in relation to open-source data gathering by the police. First, the international legal context for gathering data from openly accessible and semi-open sources, including the issue of cross-border gathering of data, is discussed. In view of the already discussed influence of national law, the paper then discusses the national legal framework for open-source investigations as laid down in the Dutch Police Act 2012 and the Code of Criminal Procedure. This analysis is used to determine if investigating open sources by the police in the Netherlands is allowed on the basis of the general task description of the police, or whether a specific legal basis and appropriate authorisation is required for such systematic observation or intelligence. Another angle covered by the paper concerns the need for open-source investigation tools to meet non-manipulability and auditing requirements associated with digital forensic quality assurance.

The second contribution by Schellekens zooms in on the relation between ethics and law, by using a specific use case of Internet robots, also known as web crawlers, spiders or simply as (ro)bots. These are used to roam the Internet and are the backbone of useful services such as search engines, auction aggregators or information gatherers for crime fighting or security. Even though proprietors of computer systems may have a legitimate interest in refusing or limiting access by robots, there are no legal prohibitions for Internet robots. This might relate to the interests of the party sending the robot, which party usually offers services that are beneficial to society as a whole. In this respect a tension exists between the interests of those that offer primary services and the interests of those that build on the primary services. This raises the question how the law should deal with such tensions? More specifically, the central problem addressed in this article is: should the law vindicate the exclusion of robots by proprietors of computers connected to the Internet? More specifically the article will examine the role of the robot exclusion protocol.

The final contribution by Koops, Hoepman and Leenes considers the challenge of embedding the identified normative and legal requirements into technical designs. The experience of the VIRTUOSO platform will be used to illustrate this strategy. Ideally, the technical development process of OSINT tools is combined with legal and ethical safeguards in such a way that the resulting products have a legally compliant design, are acceptable within society (social embedding), and at the same time meet in a sufficiently flexible way the varying requirements of different end-user groups. This paper uses the analytic framework of privacy design principles; minimise, separate, aggregate, hide, inform, control, and enforce. The paper discusses two promising approaches, revocable privacy and policy enforcement language. The approaches are tested against three requirements that seem particularly suitable for a 'compliance by design' approach in OSINT: purpose specification; collection and use limitation and data minimization; and data quality (up-to-dateness). For each requirement, the paper analyses whether and to what extent the approach could work to build in the requirement in the system. The paper demonstrates that even though not all legal requirements can be embedded fully in OSINT systems, it is possible to embed functionalities that facilitate compliance in allowing end-users to determine to what extent they adopt a 'privacy by design' approach when procuring an OSINT platform, extending it with plug-ins, and fine-tuning it to their needs. Therefore, developers of OSINT platforms and networks have a responsibility to make sure that end-users are enabled to use privacy by design, by allowing functionalities such as revocable privacy and a policy enforcement language.

All three contributions provide insights to further the knowledge on the efficient use of OSINT in a manner compatible with normative principles and legal requirements. I hope this special issue will inspire readers to contribute to the on-going debate regarding how to best explore and exploit open source information.

[27] All available from www.virtuoso.eu (accessed 15 July 2013).

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° FP7-SEC-GA-2009-242352. I want to thank Géraud Canet for leading the VIRTUOSO project, and Gabriela Bodea for doing a wonderful job in leading the legal and ethical work package within the VIRTUOSO project.

REFERENCES

Casey Eoghan. Network traffic as a source of evidence: tool strengths, weaknesses, and future needs. Dig Invest February 2004;1(1):28–43.

Casey Eoghan, editor. Digital evidence and computer crime. Forensic science, computers and the Internet. 3rd ed. Elsevier Academic Press; 2011.

Glassman Michael, Kang Min Ju. Intelligence in the internet age: the emergence and evolution of Open Source Intelligence (OSINT). Comp Hum Behav March 2012;28(2):673–82.

Koops Bert-Jaap. Colette Cuijpers en Maurice Schellekens, D3.2 Analysis of the legal and ethical framework in open source intelligence. Available at, http://www.virtuoso.eu; 1 December 2011.

Please cite this article in press as: Cuijpers C, Legal aspects of open source intelligence – Results of the VIRTUOSO project, Computer Law & Security Review (2013), http://dx.doi.org/10.1016/j.clsr.2013.09.002

Mercado Stephen. A venerable source in a new era: sailing the sea of OSINT in the information age. In: Andrew Christopher, Aldrich Richard J, Wark Wesley K, editors. Secret intelligence: A reader. Routledge; 2009. p. 78–90.

Dr. Colette Cuijpers (*cuijpers@tilburguniversity.edu*) *Assistant Professor at TILT – Tilburg Institute for Law, Technology, and Society, Tilburg University, The Netherlands.*

Colette Cuijpers Q2
TILT – Tilburg Institute for Law, Technology, and Society, Tilburg University, The Netherlands
E-mail address: cuijpers@tilburguniversity.edu

0267-3649/$ – see front matter

http://dx.doi.org/10.1016/j.clsr.2013.09.002